

Análisis sintáctico de textos espontáneos. Ensayo metodológico.

Consuelo Rodríguez Magro
Celia Villar Rodríguez
Centro de Investigación UAM-IBM

1.0. Introducción

Cuando, desde nuestro interés por los problemas del lenguaje, empezamos a hacer las primeras incursiones en este campo, pudimos comprobar la casi nula presencia de la lingüística española en él. A pesar del reconocimiento de todos los profesionales preocupados por el tema de que el lenguaje es uno de los factores más claramente afectados en cualquier patología neurológica central o periférica, hasta ahora no se han hecho estudios suficientes de análisis de la normalidad lingüística que sirvan de modelo comparativo.

Las pruebas que se utilizan para medir el lenguaje en sujetos afectados por alguno de estos síndromes (Illinois, Luria...), en todo caso, psicolingüísticos y para su interpretación es necesaria la colaboración más de un psicólogo que de un lingüista, quizá por la creencia de que un lingüista no puede interpretar un síntoma psicológico -como en efecto es- y, sin embargo, un psicólogo sí es capaz de evaluar el lenguaje de un individuo, eliminando de esta forma todo el valor científico del lenguaje y convirtiéndolo en algo cotidiano y familiar que «todos conocemos».

Así, nos encontramos en España con una total ausencia de tests puramente lingüísticos que se hubieran ocupado primero de analizar la sintaxis en discursos espontáneos tanto orales como escritos.

1.1. Hipótesis y fundamentos del trabajo

En el año 1981, en la Universidad Autónoma de Madrid, se ideó un proyecto de Doctorado conjunto entre las Facultades de Medicina y Filosofía y Letras sobre Problemas del Lenguaje, y en este contexto empezó a verse clara la necesidad de colaboración de los lingüistas en este campo; parecía haberse abierto una nueva línea de trabajo para nosotros puesto que las teorías eran fácilmente conjugables, pero la práctica estaba por hacer. La sede del Doctorado fue el Gran Hospital del Estado y allí, en el seno de una

investigación sobre la dislexia, los estudios concretos sobre muestras de población se llevaban a cabo con pruebas psicológicas, médicas, pedagógicas, económico-sociales, pero no puramente lingüísticas. Se analizaba el lenguaje en niveles primarios de la escolarización y desde pruebas pedagógicas, pero nunca, como ya hemos dicho, lingüísticas.

1.2. *Técnicas y métodos empleados*

En el marco del trabajo que se llevaba a cabo en el Gran Hospital del Estado surgió la necesidad de elaboración de una prueba que analizara la expresión espontánea (oral y escrita) y más concretamente la sintaxis.

Para ello se utilizó como punto de partida una prueba que desarrolla el análisis factorial de variables de Wagenaar, Snow y Prins¹ concebido originariamente para el estudio de las afasias y que presentaba la ventaja de admitir la cuantificación de los aspectos lingüísticos del discurso que nos interesaban. Pretendimos comprobar si las variables ofrecidas en esta prueba, eran válidas y aplicables a una muestra y un test elaborado exclusivamente para el estudio de la evolución sintáctica de niños de unas determinadas características².

1.3. *Procesamiento de los datos*

Una vez obtenidos todos los textos recogidos se pasó al análisis de los mismos a través de las variables descritas (ver nota (2)). Los datos que íbamos obteniendo completaban plantillas individuales: dos para cada niño, una para los resultados de su prueba oral y otra para los de su prueba escrita; con una sencilla multiplicación (312 por 2) nos encontramos con un total de 624 plantillas, cada una de ellas con 28 ó 30 variables, según se trate de la prueba oral o de la escrita³. El tratamiento de esta ingente cifra de datos, se realizó más cómodamente de lo que habíamos pensado en un principio, gracias a la ayuda que nos prestaron en el Centro Científico de IBM-UAM. Así pues, los programas estadísticos y la obtención de los resultados se consiguieron gracias al ordenador.

2.0. *Resultados*

2.1. *Resultados globales por colegios*

Prueba oral:

En TEMPO DEL HABLA, NOMBRES, PALABRAS CON CONTENIDO y PALABRAS FUNCIONALES, las medias van, de menor a mayor, desde el colegio 1 («Orcasitas») hasta el 4 («Estudio»), y con el orden inverso, como es lógico, en la variable SEGUNDOS DE SILENCIO, es decir, que los niños que más hablan son los que menos segundos de silencio tienen. Una salvedad

con respecto a dos de estas variables: en los colegios 2 y 4, los niños producen más palabras con contenido que funcionales, y en los otros dos no.

El colegio 1 es el que más secuencias, más cortas y menos complejas produce, en cambio el colegio «Estudio» es el que menos secuencias produce, más largas y más complejas. Pero entre los colegios 2 («Luis Vives») y 3 («Jámer») no hay unas diferencias tan claras: el «Jámer» produce más secuencias, más largas, pero menos complejas que el «Luis Vives».

Prueba escrita:

El mejor colegio en esta prueba es el «Jámer» con las medias de TEMPO, COMPLEJIDAD SINTACTICA, LONGITUD MEDIA DE LAS SECUENCIAS, PALABRAS CON CONTENIDO, PALABRAS FUNCIONALES y NOMBRES superiores a las de los demás colegios, con menos cantidad de secuencias y una media de FALTAS DE ORTOGRAFIA no muy alta.

El siguiente colegio en la media de TEMPO es el «Luis Vives», con la COMPLEJIDAD SINTACTICA y la LONGITUD MEDIA DE LAS SECUENCIAS bajas y la mayor PRODUCCION DE SECUENCIAS. Las palabras con contenido, funcionales y los nombres están en función del TEMPO, por lo tanto también son de los más altos. Y los niños de este colegio tienen más faltas de ortografía que los del resto de los colegios.

Finalmente, volviendo a comparar los dos colegios más distantes, los resultados son claramente superiores en el «Estudio» en todas las variables salvo en FALTAS DE ORTOGRAFIA, ya que los niños de este colegio tienen una media superior a la de los niños de «Orcasitas».

En ACENTUACION el mejor resultado aparece en el «Jámer» (colegio 3), seguido del «Estudio», el «Luis Vives» y «Orcasitas». Y en PUNTUACION los mejores son los del «Estudio», después los del «Jámer», «Orcasitas» y «Luis Vives», por este orden.

Observamos que en la prueba escrita sí se emplean, en todos los colegios, más palabras con contenido que funcionales.

2.2. Resultados globales por sexos

Prueba oral:

En todas estas variables aparece siempre un resultado bastante uniforme: siempre las medias de las chicas son superiores a las de los chicos, salvo en SEGUNDOS DE SILENCIO que, naturalmente es inferior. Puede parecer chocante, si pensamos en la correlación entre las variables que hemos ido comentando, que las niñas produzcan más secuencias y más largas que los

chicos, pero no resulta tan extraño si se comprueba que la diferencia entre las medias de unos y otras en la variable LONGITUD MEDIA DE LAS SECUENCIAS es mínima: 30.67 para las chicas y 29.36 para los chicos.

Prueba escrita:

El sexo femenino vuelve a presentar los mejores resultados, con medias superiores siempre. Se repite nuevamente el caso de la variable LONGITUD MEDIA DE LAS SECUENCIAS, que también es ligeramente superior (13.71 en las chicas y 13.37 en los chicos) a la de los chicos, diferencia que se ve compensada por el mayor número de palabras escrito en seis minutos. De acuerdo con un mejor nivel de escritura, las niñas tienen menos faltas de ortografía y mejores resultados tanto en la acentuación como en la puntuación.

La conclusión general de este primer análisis fue clara: que los resultados eran excesivamente dispersos. Había muchas variables con la desviación típica más alta que la media, lo cual nos lleva a pensar en una reelaboración de unas variables que resultaban demasiado discretas para la muestra elegida.

Así pues, en una segunda etapa de la investigación, por una parte los datos fueron ponderados, es decir, agrupados según el concepto que median en el discurso, ya que, en muchos casos no interesaba guardar diferencias nada significativas; y por otra parte, algunas variables desaparecieron, por no ofrecer resultados homogéneos válidos para el estudio de la sintaxis -propósito primordial de esta investigación-, o se vieron asumidas por otras, que las englobaban. De esta forma la plantilla quedó mucho más unificada con la esperanza de poder obtener datos más coherentes y útiles para nuestro estudio.

Sin embargo, los segundos resultados tampoco fueron los esperados aunque sí se consiguió una mayor homogeneidad en ellos.

2.3. Conclusiones

A pesar de ello a lo largo de todo el trabajo hemos podido llegar a algunas conclusiones generales tanto con respecto a los datos propiamente dichos como en cuanto a su relación con el medio socioeconómico en el que se producen:

Si al analizar caso por caso, los datos varían dentro del mismo grupo estudiado, esto es, dentro del mismo curso y colegio, la comparación entre distintos grupos socialmente lejanos muestra diferencias mucho más grandes sobre todo entre los extremos, en nuestro caso entre segundo curso y sexto y entre los colegios de «Orcasitas» (recuérdese que se analizaron dos colegios en este barrio madrileño) y el de

«Pozuelo». Estas diferencias se reflejan en el lenguaje, como mecanismo primordial de su comunicación, en muchas de sus características (morfológicas, léxicas, fónicas, sintácticas). Aquí se han estudiado principalmente las sintácticas.

Este trabajo corrobora con datos que el individuo no es, lingüísticamente hablando, una unidad. Un niño aislado no reacciona de la forma esperada ante las correlaciones intrínsecas que parecen tener, teóricamente y a priori, las variables. Globalmente sí hemos podido establecer esas relaciones la mayoría de las veces, pero si nos fijamos en el caso particular y estudiamos niño a niño, las diferencias y las combinaciones en el modo de comportamiento son casi tantas como el número de casos presentados. Podríamos concluir, pues, que el nivel de aprendizaje no depende tanto de los maestros como de los niños; cada uno, con sus características personales, entiende, asimila y desarrolla los mismos conceptos de manera muy diferente.

Las clases económicamente más bajas son menos homogéneas que las superiores. Los colegios «Jámer» y «Estudio» presentan unos resultados en los distintos cursos y variables más lineales que los otros colegios.

La uniformidad de las variables en los colegios «Jámer» y «Estudio» da cuenta de una uniformidad en la que intervienen tanto el ambiente escolar como el familiar, que, en este aspecto, se refuerzan recíprocamente.

Lo normal en la clase más desfavorecida es la abundancia de textos por debajo de la calidad media, mientras que, al contrario, en la clase más elevada socio-económicamente, es normal la abundancia de textos por encima de la media. Por lo tanto, no se niega la aparición de textos lingüísticamente «buenos» y «malos» en ambas clases sociales.

Después de los dos intentos de aplicación de la prueba de Wagenaar, Snow y Prins para afasias, en el análisis sintáctico de textos espontáneos, queda demostrado que la sintaxis no puede medirse con variables tan discretas como las ofrecidas en la prueba original, ni siquiera con la supresión de algunas de ellas. Esto es debido a que en el habla o en la escritura espontánea de niños normales no se producen fenómenos tan puntuales como en las de individuos afásicos o en niños con otras patologías del lenguaje, para los que fueron diseñadas las variables originales de las plantillas utilizadas. En éstos, las características anormales destacan mucho del contexto en el que aparecen y así, hallar su porcentaje con respecto a él, resulta significativo. Sin embargo,

en los niños normales estas diferencias no son tan claras, por ello quedan diluidas entre el resto de los datos.

Para obtener un nivel de normalidad en cada curso y colegio, hubiéramos necesitado triplicar - o incluso centuplicar - la muestra escogida. Pero esta ampliación tenía graves inconvenientes contra la eficacia y la operatividad de la prueba, por el tiempo que habría supuesto la transcripción y el análisis del ingente número de datos. Por otro lado, el número de 13 niños y 13 niñas por clase no fue aleatorio, sino exigido por el número aproximado de niños, no repetidores y de la edad correspondiente a su curso, que asisten a una clase cualquiera⁴.

Aumentar bastante esta cifra, hubiera significado tener que trabajar con colegios de zonas parecidas pero no iguales, con el consiguiente perjuicio para la homogeneidad de los datos recogidos.

Otro gran inconveniente viene dado por la validez de los datos. Una vez obtenidos, no serían vigentes más que unos pocos años - cada vez menos -, ya que, como sabemos, la lengua es algo vivo y por lo tanto, en constante cambio, sobre todo en estos niveles de escolaridad en los que se asume inmediatamente cualquier innovación lingüística⁵.

No obstante, estas consideraciones no nos llevan a desechar este tipo de análisis, sino que defenderemos su existencia tanto para el caso de individuos normales como para el de individuos con patologías.

Por otro lado, creemos en la necesidad de elaboración de trabajos puramente prácticos que sirvan de soporte a toda la teoría existente tanto en éste como en otros campos, así como en la conveniencia de que se lleven a cabo trabajos de conjunto, que reflejen el esfuerzo de una serie de profesionales hacia el mismo fin. Somos firmemente partidarios de los equipos multiprofesionales.

NOTAS

¹ Cf. Wagenaar, Snow y Prins (1975, 1980)

² La muestra elegida y las tablas de variables utilizadas pueden encontrarse en «Contribución lingüística al diagnóstico de la dislexia», *Actas del I Congreso Nacional de AEsLA*, Murcia, 1983. En aquella ocasión presentamos el proyecto de este trabajo; en el presente artículo ofrecemos los resultados y las conclusiones del mismo, una vez acabado.

³ Recordemos, para aquellas personas que no asistieron a dicho Congreso o para las que no puedan acceder a las Actas del mismo, que los colegios estudiados eran cuatro y no fueron elegidos al azar. Se trata de los colegios «Estudio» (clase alta), «Jámer» (clase media-alta), «Luis Vives» (clase media-baja) y «Ntra. Señora de Montserrat» (clase baja). El cupo de niños de este último grupo se completó con el colegio «Virgen de Africa» de la misma zona.

⁴ Como ya se ha dicho, en algún caso hubo que completar el cupo de niños con los de otro colegio de la zona, por falta de los suficientes en los colegios originalmente escogidos.

⁵ Este factor influye en todas las pruebas existentes y que se aplican para diferentes estudios. Los resultados obtenidos de la aplicación de cualquier test deberán ser siempre relativos, de acuerdo con la época en la que se fabricó la prueba y con las condiciones reinantes, pero en ningún caso absolutas.

BIBLIOGRAFIA

- ALONSO, A. y HENRIQUEZ UREÑA, P.: *Gramática castellana*, 2 vols., Buenos Aires, Losada, 1964.
- BERNSTEIN, B.: *Langage et classes sociales (Codes sociolinguistiques et controle social)*, París, Les Éditions de Minuit, 1975.
- BOLINGER, J., Emonds, J. et al.: *Lingüística y sociedad*. Madrid, Siglo XXI, 1976.
- BRONCKART, J.P.: *Teorías del lenguaje*. Barcelona, Herder, 1980.
- DUCROT, O. y TODOROV, T.: «Sociolingüística», en *Diccionario enciclopédico de las ciencias del lenguaje*, México, Siglo XXI, 1980, pp. 79-85.
- FISHMAN, J.A.: *Sociología del lenguaje*, Madrid, Cátedra, 1979
- GARCIA MESEGUER, A.: «El lenguaje y los sexos», *Actas del I Congreso Nacional de Lingüística Aplicada*, 1984, pp.237-252
- GUMPERZ, J. y BENNETT, A.: *Lenguaje y cultura*, Barcelona, Anagrama, 1981.
- HÖRMANN, HANS: *Psicología del lenguaje*, Madrid, Gredos, 1973.
- HUDSON, R.A.: *Sociolingüística*, Barcelona, Anagrama, 1981
- JAKOBSON, R. y HALLE, M.: *Fundamentos del lenguaje*, Madrid, Ciencia Nueva, 1967
- LABOV, W.: *Sociolinguistic patterns*, Philadelphia, University of Pennsylvania Press, 1972
- LABOV, W.: *Sociolinguistique* (Trad. del inglés por Alain Kihm, primera ed.), París, Les Éditions de Minuit, 1976
- LENNEBERG y otros: *Nuevas direcciones en el estudio del lenguaje*. Madrid, Selecta de *Revista de Occidente*, 1974
- LURIA, A. R.: «Principales formas de alocución verbal, lenguaje oral y lenguaje escrito», en *Conciencia y lenguaje*, Visor, 1984, pp. 184-187
- MALMBERG, B.: «Métodos estadísticos y matemáticos en lingüística La teoría de la información», en *Los nuevos caminos de la lingüística*, México, Siglo

XXI, 1981, pp. 206-224

MARCELLESI, J.B. y GARDIN, B.: *Introducción a la sociolingüística (La lingüística social)*, Madrid, Gredos, 1975

MARCOS MARIN, F.: *Curso de gramática española*, Madrid, Cincel-Kapelusz, 1980

MARCOS MARIN, F., RODRIGUEZ MAGRO, C. y VILLAR RODRIGUEZ, C.: «Contribución lingüística al diagnóstico de la dislexia», ACTAS DEL PRIMER CONGRESO NACIONAL DE LINGÜSTICA APLICADA, 1984, pp. 321-335

OLÉRON, P.: *El niño y la adquisición del lenguaje*, Madrid, Morata, 1981

PEÑA CASANOVA, J. y PREZ PAMIES, M.: *Rehabilitación de la afasia y trastornos asociados*, Masson, Barcelona, 1983

RODRIGUEZ MAGRO, C.: *Análisis lingüístico de grupos escolares. Ensayo metodológico*. Memoria de Licenciatura. Facultad de Filosofía y Letras, Universidad Autónoma de Madrid, 1985

SIGUAN, M.: *Lenguaje y clase social en la infancia*, ICE de la Universidad de Barcelona, Pablo del Río Editor, 1979

SLOBIN, D.I.: *Introducción a la psicolingüística*, Buenos Aires, Paidós, 1974

URIBE VILLEGAS, O.: *Sociolingüística*, México, Universidad Nacional Autónoma de México, 1970

VARIOS.: *Lecturas de sociolingüística*, Madrid, EDAF Universitaria, 1979

VILLAR RODRIGUEZ, C.: *Prueba lingüística para el análisis sintáctico de textos espontáneos. Estudio comparado de grupos escolares*. Memoria de Licenciatura. Facultad de Filosofía y Letras, Universidad Autónoma de Madrid, 1985

WAGENAAR, SNOW y PRINS: «Spontaneous speech of aphasic patients: a psycholinguistics analysis», en *Brain and language*, 2, 1975, p. 281, y en *Aphasia assessment and treatment*, edit. por M. Taylor Sarno y O. Hook, Estocolmo, Almqvist and Wiksell International, 1980, p.234.